EWD831 - 0

Why numbering should start at zero

To denote the subsequence of natural numbers 2, 3, ..., 12 without the pernicious three dots, four conventions are open to us:

a) $2 \leq i < 13$
b) $1 < i \leq 12$
c) $2 \leq i \leq 12$
d) $1 < i < 13$

Are there reasons to prefer one convention to the other? Yes, there are. The observation that conventions a) and b) have the advantage that the difference between the bounds as mentioned equals the length of the subsequence is valid. So is the observation that, as a consequence, in either convention two subsequences are adjacent means that the upper bound of the one equals the lower bound of the other. Valid as these observations are, they don't enable us to choose between a) and b); so let us start afresh.

There is a smallest natural number. Exclusion of the lower bound —as in b) and d)— forces for a subsequence starting at the smallest natural number the lower bound as mentioned into the realm of the unnatural numbers. That is ugly, so for the lower bound we prefer $\leq$ as in a) and c). Consider now the subsequences starting at the smallest natural number: inclusion of the upper bound would then force the latter to be unnatural by the time the sequence has shrunk to the empty one. That is ugly, so for the upper bound we prefer $<$ as in a) and d). We conclude that convention a) is to be preferred.

Remark. The programming language Mesa, developed at Xerox PARC, has special notations for intervals of integers in all four conventions. Extensive experience with Mesa has shown that the use of the other three conventions has been a constant source of clumsiness and mistakes, and on account of that experience Mesa programmers are now strongly advised not to use the latter three available features. I mention this experimental evidence —for what it is worth— because some people feel uncomfortable with conclusions that have not been confirmed in practice. (End of Remark.)

* * *

When dealing with a sequence of length N, the elements of which we wish to distinguish by subscript, the next vexing question is what subscript value to assign to its starting element. Adhering to convention a) yields, when starting with subscript 1, the subscript range $1 \le i < N+1$; starting with 0, however, gives the nicer range $0 \le i < N$. So let us let our ordinals start at zero: an element's ordinal (subscript) equals the number of elements preceding it in the sequence. And the moral of the story is that we had better regard —after all those centuries!— zero as a most natural number.

Remark. Many programming languages have been designed without due attention to this detail. In FORTRAN, subscripts always start at 1; in AL-

GOL 60 and in PASCAL, convention c) has been adopted; the more recent SASL has fallen back on the FORTRAN convention: a sequence in SASL is at the same time a function on the positive integers. Pity! (End of Remark.)

* * *

The above has been triggered by a recent incident, when, in an emotional outburst, one of my mathematical colleagues at the University —not a computing scientist— accused a number of younger computing scientists of "pedantry" because —as they do by habit— they started numbering at zero. He took consciously adopting the most sensible convention as a provocation. (Also the "End of...." convention is viewed as provocative; but the convention is useful: I know of a student who almost failed at an examination by the tacit assumption that the questions ended at the bottom of the first page.) I think Antony Jay is right when he states: "In corporate religions as in others, the heretic must be cast out not because of the probability that he is wrong but because of the possibility that he is right."

Plataanstraat 5
5671 AL NUENEN
The Netherlands

11 August 1982
prof. dr. Edsger W. Dijkstra
Burroughs Research Fellow